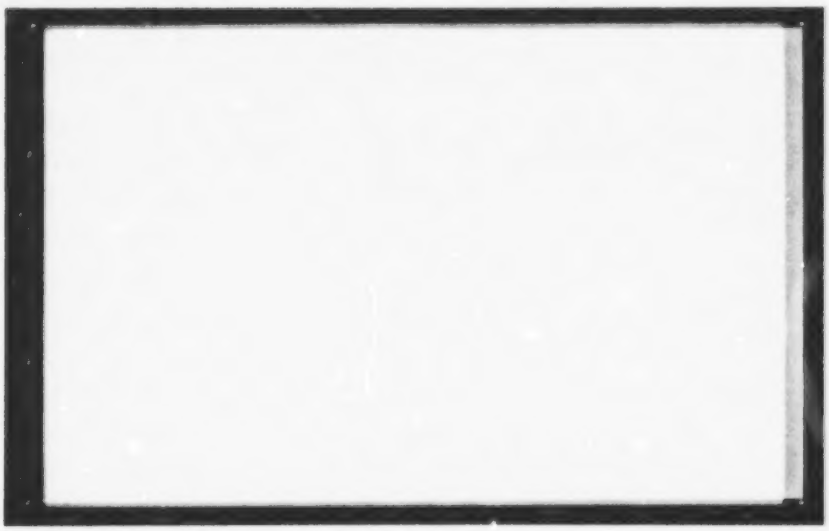
UNIVERSITÉ DE MONTRÉAL



CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

A MODIFIED NEWTON METHOD FOR SOLVING VARIATIONAL INEQUALITIES

by

Jean-Pierre Dussault* and Patrice Marcotte**

This research was supported by NSERC A-5789 and F.C.A.C. #83-AS-0026

* Département de mathématiques, Université de Sherbrooke
Sherbrooke (Québec) CANADA J1K 2R1

** GÉRAD, École des Hautes Études Commerciales,
5255, avenue Decelles, Montréal (Québec) CANADA H3T 1V6

Université de Montréal

Centre de recherche sur les transports - Publication #379
École des Hautes Etudes Commerciales - Cahier du GERAD # 85-04
Février 1985

Dépot légal : 1er trimestre 1985

ABSTRACT

Applied to strongly monotone variational inequality problems, Newton's algorithm achieves local quadratic convergence. However, global convergence cannot be guaranteed by monitoring the decrease of some convex objective function. In this paper we show how the basic Newton method can be modified to yield an algorithm that monotonically decreases the so-called "gap function" associated with the variational inequality by solving a sequence of linear programs. Convergence of the algorithm does not depend on strict or strong monotonicity assumptions; however, under strong monotonicity and strict complementarity assumptions, the set of active constraints at the solution is implicitly identified, and quadratic convergence is achieved.

Keywords : variational inequalities, nonlinear complementarity, Newton's method.

RÉSUMÉ

Il est bien connu que l'algorithme de Newton pour la résolution d'inéquations variationnelles fortement monotones possède, localement, un taux de convergence quadratique.

Dans cet article nous proposons une modification de l'algorithme de base qui soit globalement convergent, tout en conservant un taux de convergence quadratique. À chaque itération, une direction de descente pour la fonction d'écart ("gap function") est obtenue en résolvant un programme linéaire. Sous l'hypothèse moins restrictive de monotonie de la fonction de coût, l'algorithme conserve ses propriétés de convergence globale. Si, de surcroît, des hypothèses de forte monotonie et stricte complémentarité sont satisfaites, alors l'ensemble des contraintes actives est identifié, implicitement, en un nombre fini d'itérations.

0. INTRODUCTION

In this paper we consider the variational inequality problem over a convex compact polyhedron.

Since this problem is equivalent to a fixed point problem involving an upper semicontinuous cost mapping, it can be solved by a simplicial or path-following approach, for which there already exists a vast amount of litterature (see Zangwill [14], Todd [12], Saigal [11]).

For large scale problems however, these algorithms tend to become inefficient, both in terms of computer memory and running time requirements. This explains the renewed interest in algorithms closely related to the procedures devised for iteratively solving systems of nonlinear equations (Ortega and Rheinboldt [8]), such as the Jacobi, Gauss-Seidel, Newton schemes (Pang and Chan [9], Josephy [5]) or projection algorithms (Bertsekas and Gafni [2], Dafermos [4]) where the cost function is approximated, at each major iteration, by a simpler function, e.g. linear, separable or symmetric. Local or global convergence of the latter methods usually requires the a priori knowledge of upper bounds on the Lipschitz constant of the Jacobian of the cost mapping, either in a neighborhood of a solution (for local convergence) or uniformly on the feasible domain (for global convergence).

Our approach is basically different; we choose to minimize directly the complementarity term (the so-called "gap function"), associated with a nonlinear complementarity formulation of the variational inequality, using a standard first-order minimization algorithm. When the cost-mapping is monotone we show that the algorithm converges towards an equilibrium solution, and finitely so if the mapping is affine. Furthermore, under strong monotonicity and strict complementarity assumptions, the algorithm implicitly identifies the active constraint set, and achieves quadratic convergence. Since, initiated sufficiently close to the solution, the algorithm generates the same sequence of iterates as Newton's algorithm, it can be viewed as a modified Newton method for solving variational inequality problems.

1. Problem definition

Let $\Phi = \{Bx \leq b\}$ where B is an $m \times n$ matrix ($m \geq n$), represent a nonempty convex compact polyhedron in $\mathbb{R}^n$, and F be a continuously differentiable mapping from Φ into $\mathbb{R}^n$, with Jacobian F'(*).

The variational inequality problem (VIP) associated with F and Φ consists in finding some vector x in Φ, called an equilibrium solution, satisfying the variational inequality (VI):

$$(x-y)^T F(x) \leq 0 \qquad (1)$$

for all y in Φ. Since the mapping $x \to \arg\min_{y\in\Phi} (x-y)^T F(x)$ is uppersemi-continuous and Φ is convex and compact, Kakutani's fixed point theorem ([6]) guarantees the existence of a solution to VIP.

If F is the gradient of a function $f: \mathbb{R}^n \to \mathbb{R}$, i.e. if the Jacobian matrices F'(x) are symmetric, then (1) is simply the first-order necessary condition corresponding to the mathematical program:

$$\min_{x\in\Phi} \; f(x) = \int_0^x F(t)dt$$

where the line integral is unambiguously defined, being path-independent.

Definition 1.

A mapping F is:

1. Monotone on Φ if

$$(x-y)^T (F(x) - F(y)) \geq 0 \qquad \forall x,y \text{ in } \Phi .$$

2. Strictly monotone on Φ if:

$$(x-y)^T (F(x) - F(y)) > 0 \qquad \forall x,y \text{ in } \Phi \ (x \neq y) .$$

(*) All vectors are column vectors, ^{T} denotes transposition,

$$F(x) = \begin{pmatrix} F_1(x) \\ \vdots \\ F_n(x) \end{pmatrix} \text{ and } F'(x) = (\nabla F_1(x) \ \ldots \ \nabla F_n(x)).$$

3. Strongly monotone on Φ if:

$$(x-y)^T (F(x) - F(y)) \geq \alpha\|x-y\|^2 \quad \forall x,y \text{ in } \Phi$$

for some positive number α.

When F is the gradient of some differentiable function f, then the various concepts of monotonicity previously defined correspond, respectively, to convexity, strict convexity, and strong (uniform) convexity of f on Φ.

For differentiable mappings we also have the following characterizations (see Auslender [1]):

1. Monotonicity on Φ: $(x-y)^T F'(x)(x-y) \geq 0 \quad \forall x,y \in \Phi$.

2. Strong monotonicity on Φ:

$$(x-y)^T F'(x)(x-y) \geq \alpha\|x-y\|^2 \quad \forall x,y \text{ in } \Phi, \text{ for some positive } \alpha.$$

Let S denote the solution set of (1). We have that S is nonempty, convex if F is monotone, and a singleton if F is strictly monotone.

Definition 2.

The gap function g associated with a VIP is defined, for x in Φ, as:

$$g(x) = \operatorname{Max}_{y \in \Phi} (x-y)^T F(x) \tag{2}$$

We obviously have $g(x) = 0$ for all x in S. Using this concept, we can formulate (1) as a nonconvex, nondifferentiable optimization problem:

$$\operatorname{Min}_{x \in \Phi} g(x) \tag{3}$$

for which a sublinearly convergent algorithm has been proposed (Marcotte [7]).

In the remaining of the paper, we will assume that F is monotone on Φ, and F' is continuous on Φ.

Finally we will make use of the following characterization of strict complementarity:

Definition 3.

Let T denote the minimal face containing S. Then strict complementarity holds for VIP if $(x-y)^T F(x) < 0$ whenever x is in S and y is in $\Phi \setminus T$.

It is easily shown that this definition is equivalent to stating that the multipliers associated with any equilibrium solution are positive.

2. A nonlinear complementarity formulation

Proposition 1.

A feasible vector x is in S if and only if there exists a vector λ such that:

$$\begin{aligned} F(x) + B^T\lambda &= 0 \\ \lambda^T(Bx-b) &= 0 \\ \lambda &\geq 0 \end{aligned} \tag{4}$$

Proof. From (1) we deduce that any x and S must be solution to the linear program:

$$\underset{y\in\Phi}{\text{Min}} \; y^TF(x) \tag{5}$$

whose optimality conditions are precisely:

$$F(x) + B^T\lambda = 0, \quad \lambda \geq 0 \qquad \text{(dual feasibility)} \tag{6a}$$

$$By \leq b \qquad \text{(primal feasibility)} \tag{6b}$$

and $$\lambda^T(By-b) = 0 \qquad \text{(complementary slackness)} \tag{6c}$$

Replacing y by x in (6b) and (6c), we get the desired result. ∎

Proposition 2.

$$g(x) = x^TF(x) + \underset{\lambda}{\text{Min}} \; b^T\lambda \quad \text{s.t. } F(x) + B^T\lambda = 0, \; \lambda \geq 0$$

Proof.

$$\begin{aligned} g(x) &= \underset{y\in\Phi}{\text{Max}} \; (x-y)^TF(x) \\ &= x^TF(x) - \underset{\text{s.t. } By \leq b}{\text{Min}} \; y^TF(x) \\ &= x^TF(x) - \underset{\text{s.t. } B^T\mu = F(x), \; \mu \leq 0}{\text{Max}} \; b^T\mu \end{aligned}$$

by linear programming duality theory

$$= x^T F(x) + \text{Min } b^T\lambda$$
$$\text{s.t. } F(x) + B^T\lambda = 0 \quad \text{after setting } \lambda = -\mu$$
$$\lambda \geq 0 \qquad \blacksquare$$

The next result shows that any stationary point of the mathematical prog m:

$$\begin{array}{ll} \text{Min} & h(x,\lambda) \triangleq x^T F(x) + b^T\lambda \\ x\in\Phi & \\ \lambda\geq 0 & \text{s.t. } F(x) + B^T\lambda = 0 \end{array} \tag{8}$$

is in S.

Proposition 3.

Let $\begin{bmatrix} x^* \\ \lambda^* \end{bmatrix}$ be a stationary point for the nonconvex mathematical programming problem (8). Then:

1. $x^{*T}F(x^*) + b^T\lambda^* = 0$

2. $x^* \in S$.

Proof. Let $\begin{bmatrix} x^* \\ \lambda^* \end{bmatrix}$ be a stationary point for (8). It is straightforward to show that x^* is also stationary for the nondifferentiable problem (3). Suppose $g(x^*)$ is positive. Then, from a result in [7] (see also proposition 4), there exists a feasible direction d such that the directional derivative $g'(x;d)$ is negative, contradicting the stationarity of x^*. Therefore one must have that $g(x^*) = 0$, and x^* is in S. The second result follows from proposition 2. $\blacksquare$

Proposition 3 ensures that any algorithm that produces a sequence converging towards a stationary point of (8) will indeed solve VIP.

3. A linear approximation algorithm

In this section we present a model algorithm for solving (8) which proceeds by successive linear approximation of both the objective and the nonlinear part of the constraints. If the mapping F is affine, the algorithm reduces to the classical Frank-Wolfe algorithm for minimizing quadratic forms over compact convex polyhedra. In the algorithm presentation, iteration indices are omitted for enhanced readability.

Algorithm A

Let $x \in \Phi$ and $\lambda \in \arg_{\mu} \min\ b^T\mu$

$$\text{s.t. } F(x) + B^T\mu = 0, \quad \mu \geq 0$$

WHILE convergence criterion not met

DO 1. FIND DESCENT DIRECTION $\begin{pmatrix} LA(x)-x \\ LA(\lambda)-\lambda \end{pmatrix}$.

Let $\begin{pmatrix} LA(x) \\ LA(\lambda) \end{pmatrix}$ be an extremal solution to the linearized problem:

$$\begin{aligned} &\underset{\substack{y\in\Phi \\ \mu\geq 0}}{\text{Min}}\ y^T[F(x) + F'(x)x] + b^T\mu \\ &\text{s.t. } F(x) + F'^T(x)(y-x) + B^T\mu = 0 \end{aligned} \tag{9}$$

2. PERFORM ARC SEARCH:

IF $g(LA(x)) \leq \frac{1}{2} g(x)$ THEN $\theta^* \leftarrow 1$

ELSE

let $\theta^* \in \arg\min_{\theta\in[0,1]} g((1-\theta)x + \theta\, LA(x))$

3. UPDATE:

$x \leftarrow (1-\theta^*)x + \theta^* LA(x)$

$\lambda \leftarrow \lambda(x)$

where $\lambda(x) \in \arg_{\mu} \min\ b^T\mu$

$$\text{s.t. } F(x) + B^T\mu = 0, \quad \mu \geq 0$$

ELIHW

▲

Some comments are in order:

1. The algorithm can be initiated with any couple $\begin{pmatrix} x \\ \lambda \end{pmatrix}$ where $x \in \Phi$ and λ is an optimal dual solution to the primal linear programming problem (5).

2. The arc search at step 2 of algorithm A corresponds to the minimization, with respect to the primal vector x, of the nondifferentiable objective g, and could therefore be viewed as a line search. Since dual vectors λ have to be computed repeatedly for different values of the primal vector x, this operation can be carried out efficiently using reoptimization techniques of linear programming.

3. It is not required, or even advisable, that the arc search be performed exactly. Any rule (Armijo's rule for instance) that guarantees a sufficient decrease of the objective could be used.

4. Convergence analysis

We first state and prove a global convergence result for algorithm A.

Proposition 4.

Any point of accumulation of a sequence generated by algorithm A is in the solution set of VIP.

Proof. We will strive to check the hypothesis of Zangwill's convergence theorem (Zangwill [13]).

We first show that the iterates generated by algorithm A lie in a compact subspace of $\mathbb{R}^{m+n}$. Since Φ is compact by hypothesis, this result will hold if $LA(\lambda)$ stays in some compact subset C of $\mathbb{R}^m$ which is independent of x and λ.

By definition we have :

$$LA(\lambda) \in \arg\min_{\mu \geq 0} b^T\mu$$
$$\text{s.t. } B^T\mu = -F(x) + F'^T(x)(LA(x)-x)$$

Let $\gamma(x) \triangleq -F(x) + F'^T(x)(LA(x)-x) \subseteq \mathbb{R}^n$

Let also $\{N_e, e=1,\ldots,p\}$ be the set of full rank square submatrices (basis) of B^T. Since $LA(\lambda)$ is extremal, one has :

$$LA(\lambda) = N_e^{-1}\gamma(x)$$

for some $e \in \{1,\ldots,p\}$.

From the continuity of F and F', we deduce that $\gamma(x)$ must lie in a compact set K, independent of x. Therefore : $LA(\lambda) \in C$, where $C \triangleq \bigcup_{e=1}^{P} N_e^{-1} K$, which is compact.

Now it is a simple matter to show that the algorithmic map $M : \Phi \to 2^{\Phi}$ (direction finding followed by arc search) induced by algorithm A is closed, thanks to the continuity of F' over Φ. Therefore, Zangwill's convergence theorem (Zangwill [3]) will apply if we can prove that, for any feasible vector $\begin{pmatrix} \bar{x} \\ \bar{\lambda} \end{pmatrix} \in S$, a vector $\begin{pmatrix} \tilde{x} \\ \tilde{\lambda} \end{pmatrix}$ obtained by applying the algorithmic map to $\begin{pmatrix} \bar{x} \\ \bar{\lambda} \end{pmatrix}$ achieves a strictly lower objective.

Consider the VIP with F replaced by its linear approximation at $\bar{x}$:

Find $x \in \Phi$ satisfying the VI:

$$(x-y)^T (F(\bar{x}) + F'^T(\bar{x})(x-\bar{x})) \leq 0 \quad \text{for all } y \text{ in } \Phi . \tag{10}$$

Denote by $\bar{g}$ the gap function associated with (10). The corresponding linear complementarity problem can be written as the positive semidefinite quadratic programming problem:

$$\underset{x,\lambda}{\text{Min}}\ \bar{h}(x,\lambda) \triangleq x^T[F(\bar{x}) + F'^T(\bar{x})(x-\bar{x})] + b^T\lambda$$

$$\text{s.t.} \quad F(\bar{x}) + F'^T(\bar{x})(x-\bar{x}) + B^T\lambda = 0$$

$$\lambda \geq 0$$

We have: $\nabla_x \bar{h}(x,\lambda)\Big|_{x=\bar{x}} = F(\bar{x}) + F'^T(\bar{x})\bar{x} = \nabla_x h(x,\lambda)\Big|_{x=\bar{x}}$, which implies that the linear approximation (Frank-Wolfe) descent direction for $\bar{h}$ at $\begin{pmatrix} \bar{x} \\ \bar{\lambda} \end{pmatrix}$ is identical with $\begin{pmatrix} LA(\bar{x})-\bar{x} \\ LA(y)-\bar{\lambda} \end{pmatrix}$

If we define $\Gamma(\bar{x})$ as :

$$\Gamma(\bar{x}) \triangleq \arg\min_{z \in \Phi} z^T F(\bar{x})$$

we obtain :

$$g'(\bar{x};d) = \max_{y\in\Gamma(\bar{x})} d^T(F(\bar{x}) + F'^T(\bar{x})(y-\bar{x}))$$
$$= \bar{g}'(\bar{x};d)$$
$$< 0 \ .$$

Consequently, minimization of g in the direction d will yield a strictly lower value, i.e.: $g(\hat{x}) < g(\bar{x})$ for all $\hat{x}\in M(\bar{x})$, as desired. ■

Remark. It would have been possible to state the algorithm without explicit reference to the dual variable λ. Indeed, at each stage of the algorithm, the variable λ is determined by x, and can be considered a dependent variable. Δ

The remainder of this section will be devoted to the study of the convergence rate of algorithm A.

Proposition 5.

If F is affine, then algorithm A converges towards a point in S in a finite number of steps.

Proof. Let $F(x) = Ax + a$. Then (8) reduces to the convex quadratic program:

$$\begin{aligned} \min_{x,\lambda} \ & x^T(Ax + a) + b^T\lambda \\ \text{s.t.} \ & Ax + a + B^T\lambda = 0 \\ & Bx \le b \\ & \lambda \ge 0 \end{aligned} \tag{11}$$

In this instance, algorithm A is equivalent, let the choice of θ at step 2, to the Frank-Wolfe algorithm.

It is well-known that there exists an extremal solution to (11). Let $u^1, u^2, \ldots, u^\ell$ be those solutions, i.e. a subset of the extreme points of $\Phi \times \mathbb{R}^{n+}$. At some iteration of algorithm A, we must have that $LA(x) = u^i$, $i\in\{1,2,\ldots,\ell\}$, otherwise all iterates would stay some distance away from S. Now the optimal stepsize will equal one, yielding an optimal solution $x^* = u^i$ in a finite number of steps. ■

In the nonlinear case, we will show that, for x sufficiently close to $x^* \in S$, the stepsize will always be equal to one, and the next iterate will be equivalent to the Newton iterate NEW(x), solution of the VI:

$$[NEW(x)-y]^T[F(x) + F'^T(x)(NEW(x)-x)] \leq 0 \quad \text{for all } y \text{ in } \Phi \tag{12}$$

In Josephy [5] it is shown that Newton's method is locally quadratically convergent, under a strong regularity condition (Robinson [10]).

We now prove that, for x sufficiently close to a solution x*, NEW(x) lies on the optimal face T, i.e. the set of active constraints at x* is implicitly identified.

From now on, it is assumed that F is strongly monotone and that strict complementarity holds at the solution (recall definitions 1 and 3).

Proposition 6.

There exists a neighborhood of the solution x* such that if a feasible point x lies in this neighborhood, then NEW(x) is in T.

Proof. If $T = \Phi$, the result holds trivially. Otherwise, there exists an infinite subsequence $\{x^k\}_{k\in I}$ of the sequence $\{x^k\}$ generated by algorithm A such that x^k always lie in the relative interior of some face $\overline{T}$ distinct from T. Let u be any extreme point of $\overline{T}$ not in T. We have:

$$x^k \underset{\substack{k\to\infty \\ k\in I}}{\to} x^*$$

$$\text{but : } [NEW(x^k)-u]^T[F(x^k) + F'^T(x^k)(NEW(x^k)-x^k)] = 0$$

for all $k\in I$, and some extreme point u of Φ ($u \notin T$). It is easy to show that $\lim_{\substack{k\to\infty \\ k\in I}} NEW(x^k) = NEW(x^*) = x^*$.

It follows, by continuity, that $(x^*-u)^T F(x^*) = 0$,
which contradicts the strict complementarity assumption. ∎

Remark. Similar results hold for all perturbation or approximation algorithms, in particular for the projection algorithm where, at a given iterate x^k, F is approximated by the linear mapping $Mx + \rho F(x^k) - Mx^k$. M is usually a positive definite symmetric matrix, and ρ denotes a positive relaxation factor. Δ

The proof of order two convergence will be based on the fact that, in a neighborhood of x*, θ will always be set to one, and we will have: LA(x) = NEW(x).

First, let us notice that NEW(x), together with a suitable multiplier vector $\lambda(NEW(x))$, solves the linearized problem:

$$\begin{aligned} \underset{y\in\Phi}{\text{Min}} \quad & y^T[F(x) + F'(x)(y-x)] + b^T\mu \\ \text{s.t.} \quad & F(x) + F'^T(x)(y-x) + B^T\mu = 0 \\ & \mu \geq 0 \end{aligned} \tag{13}$$

Let $\psi(x) \triangleq \{\binom{y}{\mu} \mid \mu \geq 0 \text{ and } F(x) + F'^T(x)(y-x) + B^T\mu = 0\}$

$\subseteq R^{m+n}$

We know that there exists an optimal solution $\begin{pmatrix} NEW(x) \\ \lambda(NEW(x)) \end{pmatrix}$ to (13) which is an extremal point of $(\Phi \times R^{m+}) \cap \psi(x)$. Actually, if $\psi(x^*)$ were known beforehand, VIP could be solved as a linearly constrained problem. Newton's method gradually improves the approximation $\psi(x^k)$ to $\psi(x^*)$. As x approaches the equilibrium solution x*, we will have LA(x) = NEW(x).

Proposition 7.

There exists a neighborhood of the solution x* such that LA(x) = NEW(x) whenever a feasible point x lies in this neighborhood.

Proof. By strict complementarity and strong monotonicity, $\begin{pmatrix} x^* \\ \lambda^* \end{pmatrix}$ must be unique. Therefore, for x close enough to x^*, there can exist but a single extremal point of $\Phi \cap \psi(x)$ (*) lying in an arbitrary small neighborhood around $\begin{pmatrix} x^* \\ \lambda^* \end{pmatrix}$; NEW(x) has to be this point, and so does $\begin{pmatrix} LA(x) \\ LA(\lambda) \end{pmatrix}$. ∎

Proposition 8.

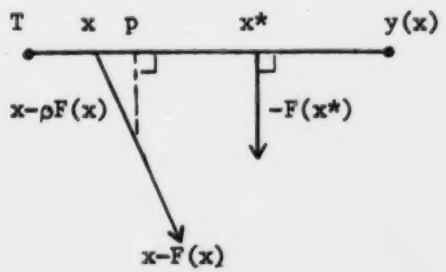
There exists a neighborhood of the solution x^* such that for any feasible x in this neighborhood, one has: $g(NEW(x)) \leq \frac{1}{2}\, g(x)$.

Proof. Since local convergence of Newton's method is quadratic, it suffices to show that $g(x) = 0\|x-x^*\|$. (**)

First case: The optimal face T is a singleton.

Let $T = \{u\} = \{x^*\}$ for some extreme point u of Φ. Then from proposition 6, we get: $g(NEW(x)) = g(NEW(u)) = 0$ and the result holds trivially.

Second case: $T \neq \{x^*\}$ and $x \in T$.



Let ρ be a positive number such that the mapping $Proj_{\Phi}(x-\rho F(x))$ is a contraction, where $Proj_{\Phi}$ denotes the projection operator on Φ in the usual euclidian norm. The existence of such a number ρ is a consequence of results to be found in Dafermos [4].

(*) Slightly abusing notation, we write $\Phi \cap \psi(x)$ instead of the notationally correct $(\Phi \times \mathbb{R}^{m+}) \cap \psi(x)$.

(**) By $g(x) = 0\|x-x^*\|$, we mean the existence of positive constants ζ_1 and ζ_2 such that $\zeta_1\|x-x^*\| \leq g(x) \leq \zeta_2\|x-x^*\|$ for all x in Φ.

For x sufficiently close to x*, we have that

$p \triangleq \text{Proj}_{\Phi}(x-\rho F(x))$ lies in T, i.e.:

$$p = \text{Proj}_{T}(x-\rho F(x))$$

Now: $\|p-x^*\| \leq \theta \|x-x^*\|$ with $\theta \in [0,1)$

Thus: $\|p-x\| \geq \|x^*-x\| - \|p-x^*\|$

$$\geq (1-\theta)\|x^*-x\| \qquad (14)$$

And: $$(x-p)^T F(x) = \frac{1}{\rho}\|x-p\|^2 \qquad (15)$$

Let $\psi \triangleq \underset{\varphi>0}{\text{Max}}\{\varphi \mid x + \varphi(p-x) \in T\}$ and $y(x) \triangleq x + \psi(p-x)$

We have: $(x-y(x))^T F(x) = \psi (x-p)^T F(x)$

$$= \frac{\psi}{\rho}\|x-p\|^2 \quad \text{by (15)}$$

$$\geq \|\psi(x-p)\| \frac{1-\theta}{\rho} \|x^*-x\| \quad \text{by (14)}$$

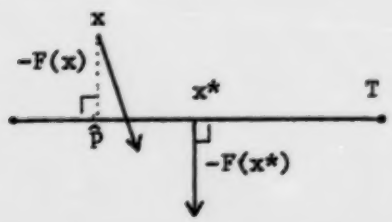
Now it is easy to show that $\|\psi(x-p)\| = 0(1)$ since x^* lies in the relative interior of T. It follows that:

$$g(x) \geq (x-y(x))^T F(x)$$

$$\geq 0\|x^*-x\|$$

Third case: $x \notin T$. Therefore: $T \neq \Phi$.

Let $\hat{p} = \text{Proj}_{T}(x)$



By strict complementarity, one has: $\cos(x-p, F(x^*)) \geq \gamma > 0$,

or: $(x-\hat{p})^T F(x^*) \geq \gamma\|x-\hat{p}\| \cdot \|F(x^*)\|$

therefore: $(x-\hat{p})^T F(x) \geq \frac{\gamma}{2}\|x-\hat{p}\| \cdot \|F(x^*)\|$ (16)

if x is sufficiently close to x*.

We consider two mutually exclusive subcases:

1. $\|x-\hat{p}\| \geq \|\hat{p}-x^*\|$

Then: $\|x-x^*\|^2 \leq 2\|x-\hat{p}\|^2$, implying: $\frac{1}{\sqrt{2}}\|x-x^*\| \leq \|x-\hat{p}\|$. (17)

And: $g(x) \geq (x-\hat{p})^T F(x)$

$$\geq \frac{\gamma}{2\sqrt{2}}\|F(x^*)\|\cdot\|x-x^*\| = 0\|x^*-x\|$$

with $F(x^*) \neq 0$ since $T \neq \Phi$.

2. $\|x-\hat{p}\| < \|\hat{p}-x^*\|$

We have: $g(x) \geq (x-y(\hat{p}))^T F(x)$

where $y(\hat{p})$ is defined similarly to $y(x)$ in the first case.

$g(x) \geq (\hat{p}-y(\hat{p}))^T F(x)$

$\geq \frac{1}{2}(\hat{p}-y(\hat{p}))^T F(\hat{p})$ for x close to x^*

$\geq 0\|\hat{p}-x^*\|$ as in the second case, p playing the role of x.

$= 0\|x-x^*\|$ Since $\|\hat{p}-x^*\| \geq \frac{1}{2}\|x-x^*\|$

Now, in all cases: $g(x) = (x-y)^T F(x)$ for some $y \in \Phi$

$= (x-x^*)^T F(x) + (x^*-y)^T F(x)$

$\leq \|x-x^*\|\cdot\|F(x)\| + 0\|x-x^*\|$

$= 0\|x-x^*\|$

Therefore: $g(x) = 0\|x-x^*\|$.

Since the convergence of Newton's method is locally quadratic, it must eventually decrease g quadratically, and the result follows. ∎

Combining the previous results we obtain:

Theorem.

Let $\{x_k\}$ be any sequence generated by algorithm A. Then we have $\lim_{k\to\infty} g(x_k) = 0$ and $g(x_{k+1}) < g(x_k)$ whenever x_k is not a solution to VIP.

Furthermore, if the mapping F is strongly monotone, and strict complementarity holds at the (unique) solution x*, then the set of active constraints of Φ is reached after a finite number of iterations, and the convergence of the sequence x_k towards x* is quadratic.

Proof. The first assertion follows from the descent nature of Algorithm A.

Now let Z_1, Z_2 and Z_3 be the neighborhoods whose existence is guaranteed by propositions 6, 7 and 8 respectively, and set $Z = Z_1 \cap Z_2 \cap Z_3$. Since algorithm A is convergent, the iterates must eventually enter and stay inside Z. From propositions 7 and 8, we have that the stepsize θ is equal to one from that point on, and this corresponds to a Newton step. Proposition 6 indicates that the current iterate (or the next) must lie in T. ∎

Remarks.

1. When x is in Z, algorithm A is equivalent to Newton's method for Solving the system of nonlinear equations

$$F(x) + \bar{B}^T\bar{\lambda} = 0$$
$$\bar{B}x = \bar{b}$$

where T = $\{x \mid \bar{B}x=\bar{b}\}$ and $\bar{\lambda}$ is the vector of non zero multipliers.

2. If the mapping F is affine and antisymmetric, i.e. $F(x) = (A-A^T)x + a$, then (8) reduces to a linear program.

3. When T is the singleton {x*} (first case in the proof of proposition 8) finite termination occurs. Δ

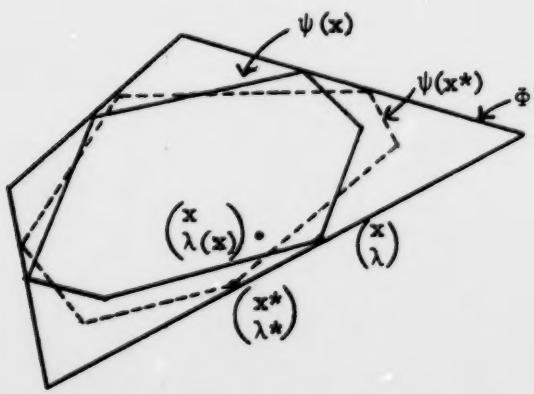


Figure 1. Illustration of algorithm A.

5. Conclusion

The main result of this paper has been to develop a globally and quadratically convergent algorithm for solving monotone VIP. The algorithm has the property of decreasing the gap function in a monotone fashion, operates in the space $\mathbb{R}^{m+n}$ of primal and dual variable (multipliers) and, under the usual strict complementarity assumption, achieves quadratic convergence. Although it is locally equivalent to Newton's algorithm, each iteration consists of solving a linear programming (rather than linear complementarity) subproblem.

REFERENCES

[1] AUSLENDER, A. (1976), Optimisation. Méthodes numériques, Masson, Paris.

[2] BERTSEKAS, D. and GAFNI, E.M. (1982), "Projection methods for variational inequalities with application to the traffic assignment problem", Math. Prog. Study 17.

[3] COTTLE, R.W. and DANTZIG, G.B. (1967), Positive (semi) definite programming, in: Nonlinear Programming, J. Abadie ed., North-Holland.

[4] DAFERMOS, S.C. (1983), "An iterative scheme for variational inequalities", Math. Programming 26.

[5] JOSEPHY, N.H. (1979),"Newton's method for generalized equations", Tech. report # 1966, Math Research Center, University of Wisconsin, Madison.

[6] KAKUTANI, S. (1941), "A generalization of Brouwer's fixed point theorem", Duke Mathematics Journal 8.

[7] MARCOTTE, P. (1983), "A new algorithm for solving variational inequalities with application to the traffic assignment problem", To appear in Mathematical Programming.

[8] ORTEGA, J.M. and RHEINBOLDT, W.C. (1970), Iterative solution of nonlinear equations in several variables, Academic Press.

[9] PANG, J.S. and CHAN, D. (1982), "Iterative methods for variational and complementarity problems", Math. Programming 24.

[10] ROBINSON, S.M. (1983), "Generalized equations". Mathematical programming: the state of the art, Grötschel, Korte eds, Springer Verlag.

[11] SAIGAL, R., "Fixed point computing methods", in Operations Research Support Methodology, Holzman, A.G. ed., Dekker.

[12] TODD, M.J. (1976), The computation of fixed point and applications, Springer.

[13] ZANGWILL, W.I. (1979), Nonlinear Programming: A Unified Approach, Prentice Hall.

[14] ZANGWILL, W.I. and GARCIA, C.B. (1981), "Equilibrium programming: the path-following approach and dynamics", Math. Programming 21.